NEW FACE DETECTION METHOD BASED ON MULTI-SCALE HISTOGRAMS

Saravanan.R
Assistant Professor
Sri Sunflower College of Engineering & Technology
Lankapalli
Andhrapradesh, India
Saranme1990@gmail.com

\

Jayanthi.S
Professor/Dean(R&D),
Samskruti College of Engineering and Technology
Kondapur
Telangana , India
nigilakash@gmail.com

Abstract—*Face detection plays a significant role in many applications, such as video surveillance, gender classification, and facial recognition. In this paper, we propose a new face detection method based on multi-scale histograms. The proposed method uses a multi-scale histogram to represent a face, thereby improving computational efficiency, and making the process suitable for big-data multimedia databases. The experimental results show that the proposed method can achieve accuracy rates similar to the LBP-based method, but at a speed over 10 times more quickly when the block size is properly set.*

Keywords: face detection; multi-scale descriptor; texture feature

I. INTRODUCTION

The technology for face detection and face recognition has advanced rapidly in recent years. Face detection, the first step in face recognition, plays an important role in the process and has been recognized as a complicated and difficult area for research. Issues related to face detection include accuracy of detection and efficiency of performance. There has been a considerable number of face detection papers, which have proposed methods that can be divided into four types [14]: 1) Template matching, 2) feature invariant, 3) knowledge-based, and 4) appearance-based.

Template matching [2, 9] approach starts with a defined manual standard sample and then searches for possible face areas through the search window. Although this method is simple and relatively easy to implement, it is easily affected by the direction, size, and rotation of the face.

Feature invariant [4, 8, 10, 11, 15, 16] methods use edge detection to extract facial features, such as eyebrows, eyes, nose, mouth and skin, and then describe the relationship between the features statistically. The disadvantage of this method is that these facial features are easily affected by brightness, noise, and other environmental factors.

Knowledge-based method [18] transforms the relationship between the general features of a face into a number of rules. For example, facial images usually include a pair of eyes, a nose and a mouth, so some rules can be set about the distance and position between them to represent a face. However, the disadvantage of the knowledge-based methods lie indetermining how to extract useful rules to match facial features. If the matching rules are not properly set, accuracy is negatively impacted.

Appearance-based methods [5, 6, 7] differ from the knowledge-based methods in that it does not have to define rules to describe faces. Generally, the method identifies features of the human face and the non-facial area through statistical analysis and machine learning. The disadvantage of this method is that it requires a number of learning samples to increase accuracy, and it is also easily affected by the quality of the learning samples.

Ahonen et. al. [1] presented a novel facial image representation based on the local binary pattern (LBP) texture feature. The LBP operator [12] is one of the best texture descriptors and has been widely used in various applications. The LBP operator was originally designed for texture description. The operator appoints a label to each pixel of an image by using the center pixel value to make the 3 × 3 neighbor pixels binary. After that, the histogram of the labels is used as a texture descriptor. Fig. 1 shows an illustration of the basic idea behind the LBP operator. The LBP-based face detection methods require many computations and are noise sensitive.

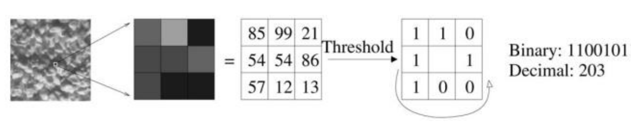

Fig. 1. The basic LBP operator.

In this paper, we propose a new face detection method based on multi-scale histograms. The proposed method uses a multi-scale histogram to represent a face, which is more noise resistant than the LBP-based method and requires less computation. The latter property makes the proposed method more suitable for a big-data multimedia system. This paper is organized as follows: Section 2 discusses the proposed method in detail; Section 3 presents the experimental results; and Section 4 provides the conclusions.

II. PROPOSED METHOD

This section presents a method for face detection that is based on a coarse-to-fine texture descriptor. Our system consists of several steps. It starts with a preprocessing step for the input facial images. Then, the images are normalized to a size of 128×128 pixels and divided into $N \times N$ non-overlapping blocks. Next, the proposed coarse-to-fine method is used to extract the features for SVM. SVM is used as a classifier to differentiate face and non-face. The system framework diagram is shown in Fig. 2.

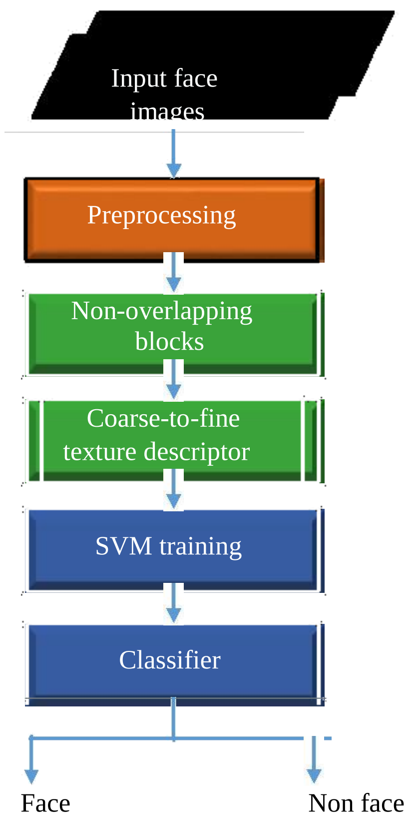

Fig. 2. Framework diagram

A. Preprocessing

As mentioned above, the first step for an image region is to normalize the image to a size of 128×128 pixels and then divide it into $N \times N$ non-overlapping blocks as shown in Fig. 2. Each block contains $m \times m$ pixels. The following steps are then performed on each color channel.

The mean of the block will be used to transform the block into binary. However, blocks are exposed to unstable 0/1 bits when the block is smooth. The threshold value will therefore be added to the mean value (m) to reduce the influence before comparison as shown in Eq. (1) [17]. The value ofġ TH_{smooth}ġ will be determined experimentally.

$$b_{ij} = \begin{cases} 0, & if \ x_{ij} < \mathrm{m} + TH_{smooth} \\ 1, & otherwise \end{cases} \quad (1)$$

B. Coarse-to-fine texture descriptor

A 1-bit representation of pixels is shown in the previous texture description of 0/1 bit. The structure of the coarse-to-fine description for the texture description can be easily derived from the 1-bit mode. If the block is not smooth, the 1-bit mode can be further extended into a 2-bit mode, generating 2 finer means, i.e., high mean (hm) and low mean (lm). The '0' region generates the lm, and the '1' region generates the hm from the 1-bit mode. The bitmap (BM) for the 2-bit mode can be generated by Eq. 2.

$$b_{ij} = \begin{cases} 11, if \ x_{ij} \geq hm + TH_{smooth}, \\ 10, if \ m + TH_{smooth} \leq x_{ij} < hm + TH_{smooth}, \\ 01, if \ lm + TH_{smooth} \leq x_{ij} < m + TH_{smooth}, \\ 00, if \ x_{ij} \leq lm + TH_{smooth}. \end{cases} \quad (2)$$

The 3-bit mode can be produced by further extending the 2-bit mode. The two means of the 2-bit mode, lm and hm, will be extended into four finer means, i.e., llm (low low mean), lhm (low high mean), hlm (high low mean), hhm (high high mean). The BM for the 3-bit mode can be generated by Eq. 3.

$$b_{ij} = \begin{cases} 111, if \ x_{ij} \geq hhm + TH_{smooth}, \\ 110, if \ hm + TH_{smooth} \leq x_{ij} < hhm + TH_{smooth}, \\ 101, if \ hlm + TH_{smooth} \leq x_{ij} < hm + TH_{smooth}, \\ 100, if \ m + TH_{smooth} \leq x_{ij} < hlm + TH_{smooth}, \\ 011, if \ lhm + TH_{smooth} \leq x_{ij} < m + TH_{smooth}, \\ 010, if \ lm + TH_{smooth} \leq x_{ij} < lhm + TH_{smooth}, \\ 001, if \ llm + TH_{smooth} \leq x_{ij} < lm + TH_{smooth}, \\ 000, if \ x_{ij} < llm + TH_{smooth}, \end{cases} \quad (3)$$

The multi-scale descriptor of the coarse-to-fine features using the binary pattern provides a crucial way to extract the texture features in the image. Blocks of the image are exposed to different complexities; therefore, there will be different k-bit modes for each blocks in our method. The 1-bit mode can be used to represent smooth blocks, and 2-bit and 3-bit modes are more suitable for distinguishing complicated blocks.

C. Descriptor Construction

In the proposed method, each block will be represented by 1-, 2- and 3-bit modes to fit the feature of the block more accurately. After the multi-scale representation is generated, the 8-bin histogram will accumulate the counts of the bit patterns in each block, which will be used as the final texture descriptor. For the 1-bit mode, because there are only two bit patterns, 0 and 1, the count of the zeros (0) will go into the first four bins and the count of the ones (1) will go into the lastfour bins in the histogram. As the example in Fig. 3 shows, if the histogram is built for the 1-bit mode, the resultant histogram will be [10, 10, 10, 10, 6, 6, 6, 6]. For the 2-bit mode, there will be four bit patterns: 00, 01, 10 and 11. Likewise, the count of the first pattern (00) will go into the first two bins, the count of the second pattern (01) will go into the next two bins, and so on. Using the 2-bit mode example in Fig. 3, the resultant histogram should be [8, 8, 2, 2, 5, 5, 1, 1]. Finally, for the 3-bit, because there are eight bit patterns, each bin of the histogram corresponds to each of the bit patterns for the 3-bit mode in Fig. 3. The resultant histogram is [6, 2, 4, 0, 3, 0, 1, 0].

80	90	100	110
85	95	105	115
90	100	110	120
95	105	115	125

0	0	0	1
0	0	0	1
0	0	1	1
0	0	1	1

00	00	00	10
00	00	01	10
00	00	10	10
00	01	10	11

000	000	001	010
000	000	010	100
000	001	010	100
000	010	100	110

Fig. 3. The process of generating binary bitmap from coarse to fine

After constructing the histogram for each block in an image, the final descriptor will be obtained by concatenating the histograms from all the blocks in the image. As an image is divided into M×M blocks, the dimension of the descriptor is M×M×24. Fig. 4 shows an example of the descriptor construction for 4×4 blocks. The dimension of the descriptor is 24576 bins (24 bins×1024 blocks). If the color image is applied, the descriptor will be computed separately for each color channel, and then the output histogram of each channel will be combined into the final histogram.

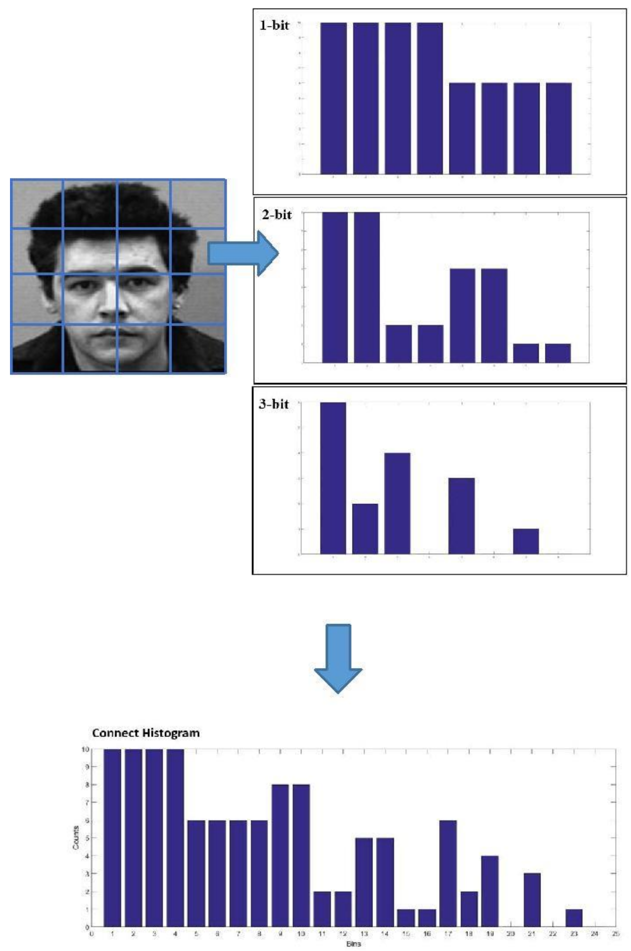

Fig. 4. Example of descriptor construction

III. EXPERIMENTS

In our experiments, the Libor Spacek's facial images databases are used [13]. For the training process, we used 60 face images for the positive samples and 90 face images for negative samples. After training samples, we applied

SVM [3] classifier to obtain the results of the training step. For the testing process, we used 2,000 images for positive samples and negative samples. Table 1 shows the precision and the average detection time for our proposed method under different block sizes. Table 2 shows the results using the LBP descriptor [1]. The results of the experiment prove that our
Proposed method is more cost effective than LBP descriptor while achieving nearly the same precision rate.

Table 1. Coarse to fine texture descriptor

Size of block	Precision for positive samples	Precision for negative samples	Detection time(ms)
4 x 4	99%	98%	3.51799
8 x 8	99%	97%	2.63614
16 x 16	96%	96%	1.22268
32 x 32	98%	91%	0.833912

Table 2. LBP texture descriptor

Size of block	Precision for positive samples	Precision for negative samples	Detection time(ms)
4 x 4	99%	98%	41.2103
8 x 8	99%	99%	11.0748
16 x 16	99%	99%	3.4101
32 x 32	98%	99%	1.51305

IV. CONCLUSION

Our paper proposes a coarse-to-fine texture-based image descriptor for face detection. We use a multi-scale histogram to represent a face, which is the main contribution of this paper. Compared with the LBP descriptor, our approach has similar precision but is more efficient and cost effective. In the future, we will investigate more properties of the multi-scale structure and use a large facial database to verify the reliability and robustness of the proposed method.

ACKNOWLEDGMENT

This work was supported by Ministry of Science and Technology, Taiwan, under Grants MOST 103-2221-E-468-007-MY2 and MOST 104-2218-E-468-001.

REFERENCES

[1] Timo Ahonen, Abdenour Hadid, Matti Pietikainen, "Face Descriptor with Local Binary Patterns: Application to Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, 2006, pp.2037-2041

[2] Ian Craw, David Tock, and Alan Bennett, "Finding Face Features," *Proc. Second European Conf. Computer Vision*, pp. 92-96, 1992.

[3] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A Library for Support Vector Machines", *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no.27, pp.1-27, 2011.

[4] Ying Dai and Yasuaki Nakano, "Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene, "*Pattern Recognition*, vol. 29, no. 6, pp. 1007-1017, 1996.

[5] Lin-Lin Huang, Akinobu Shimizu, Yoshihiro Hagihara, Hidefumi Kobatake, "Face Detection from Cluttered Images Using a Polynomial Neural Network", *Neurocomputting* vol. 51, 2003, pp. 197-211.

[6] Lin-Lin Huang, Akinobu Shimizu, Yoshihiro Hagihara, Hidefumi Kobatake, "Gradient Feature Extraction for Classification-Based Face Detection", *Pattern Recognition* vol. 36, issue 11, 2003, pp. 2501-2511.

[7] Lin-Lin Huang, Akinobu Shimizu, Hidefumi Kobatake, "Robust Face Detection Using Gabor Filter Features", *Pattern Recognition Letters*, vol. 26, issue 11, 2005, pp. 1641-1649.

[8] R. Kjeldsen and J. Kender, "Finding Skin in Color Images, "*Proc. Second Int'l Conf. Automatic Face and Gesture Recognition*, pp. 312-317, 1996_

[9] A. Lanitis, C. J. Taylor, and T. F. Cootes, "An Automatic Face Identification System Using Flexible Appearance Models," *Image and Vision Computing*, vol. 13, no. 5, pp. 393-401, 1995.

[10]T. K. Leung, M. C. Burl, and P. Perona, "Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching, "*Proc. Fifth IEEE Int'l Conf. Computer Vision*, pp. 637-644, 1995.

[11]S. McKenna, S. Gong, and Y. Raja, "Modeling Facial Color and Identity with Gaussian Mixtures, "*Pattern Recognition*, vol. 31, no. 12, pp. 1883-1892, 1998.

[12]T. Ojala, M. Pietika¨inen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions", Pattern Recognition, vol. 29, no. 1, pp. 51-59, 1996.

[13]Dr Libor Spacek Facial Images Databases http://cswww.essex.ac.uk/mv/allfaces/faces94.html.

[14]Ming-Hsuang Yang, David J. Kriegman, narendra Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, 2002, pp. 34-58.

[15] K. C. Yow and R. Cipolla, "Feature-Based Human Face Detection, "*Image and Vision Computing*, vol. 15, no. 9, pp. 713-735, 1997. Jie Yang and Alex Waibel, "A Real-Time Face Tracker," *Proc. Third Workshop Applications of Computer Vision*, pp. 142-147, 1996.

[16] Chia-Hung Yeh, Chih-Yang Lin, Kahlil Muchtar, and Li-Wei Kang, "Real-time background modeling based on a multi-level texture description", *Information Sciences*, vol. 269, no. 10, pp.106-127, 2014.

[17] Tiping Zhang, Yuan Yan Tang, Bin Fang, Zhaowei Shang, Xiaoyu Liu, "Face Recognition Under Varying Illumination Using Gradient faces", *IEEE Transactions on Image Processing*, vol. 18, 2009, pp. 2599-2606.

Mr.R.Saravanan was born in the year 1990.He received his Bachelor degree in Information Technology from Anna University, India in 2011, and Master degree in Computer Science and Engineering from Anna University, in 2014, respectively. He has 4 years of teaching experience. His area of research includes Data Mining, Neural Networks and Big Data analysis. So far he has published 4 papers in various national, International conferences and journals.

Dr.S.Jayanthi was born in the year 1981. She received her Bachelor degree in Computer Science from Bharathidasan University, India in 2002, and Master degrees in Computer Applications, and Computer Science and Engineering from Bharathidasan University, in 2005, and from Anna University of Technology, India in 2009, respectively. She obtained her Ph.D from Karpagam University, Coimbatore, India. She has 8 years of teaching experience. Her area of research includes Data Mining, Neural Networks and Big Data analysis. So far she has published 26 papers in various national, International conferences S. and journals.